A category-specific advantage for numbers in verbal short-term memory: Evidence from semantic dementia

Elizabeth Jefferies [a,*], Karalyn Patterson [b], Roy W. Jones [c], David Bateman [d], Matthew A. Lambon Ralph [e]

[a] *University of Bristol, Bristol, UK*
[b] *MRC Cognition and Brain Sciences Unit, Cambridge, UK*
[c] *St. Martin's Hospital, Bath, UK*
[d] *Royal United Hospital, Bath, UK*
[e] *University of Manchester, Manchester, UK*

Abstract

This study explored possible reasons for the striking difference between digit span and word span in patients with semantic dementia. Immediate serial recall (ISR) of number and non-number words was examined in four patients. For every case, the recall of single-digit numbers was normal whereas the recall of non-number words was impaired relative to controls. This difference extended to multi-digit numbers, and remained even when frequency, imageability, word length, set size and size of semantic category were matched for the numbers and words. The advantage for number words also applied to the patients' reading performance. Previous studies have suggested that semantic memory plays a critical role in verbal short-term memory (STM) and reading: patients with semantic dementia show superior recall and reading of words that are still relatively well known compared to previously known but now semantically degraded words. Additional assessments suggested that this semantic locus was the basis of the patients' category-specific advantage for numbers. Comprehension was considerably better for number than non-number words. Number knowledge may be relatively preserved in semantic dementia because the cortical atrophy underlying the condition typically spares the areas of the parietal lobes thought to be crucial in numerical cognition but involves the inferolateral temporal-lobes known to support general conceptual knowledge.
© 2003 Elsevier Ltd. All rights reserved.

Keywords: Immediate serial recall; Matching span; Number knowledge; Semantic dementia

1. Introduction

Immediate serial recall (ISR) tasks, like digit and word span, are traditionally considered to measure the functioning of a discrete phonological short-term memory (STM) system. It is now commonly acknowledged, however, that verbal ISR draws on multiple levels of representations that play a role in language production and comprehension. Normal participants show effects of lexicality, word frequency, imageability, concreteness and semantic blocking in ISR (Bourassa & Besner, 1994; Hulme, Maughan, & Brown, 1991; Hulme et al., 1997; Poirier & Saint Aubin, 1995; Walker & Hulme, 1999), suggesting that lexical and seman-

tic representations play a role in the maintenance of phonological information over brief periods of time.

Some complementary evidence is provided by neuropsychology (see Martin & Lesch, 1996; Martin & Saffran, 1997; Martin, Shelton, & Yaffee, 1994). This study examines ISR in patients with semantic dementia (SD), who show a specific decline in semantic memory. SD is the temporal-lobe variant of frontal–temporal dementia, and is associated with progressive focal atrophy of the inferolateral–temporal neocortex. SD patients are anomic and have impaired comprehension on a range of verbal and non-verbal tasks (Bozeat, Lambon Ralph, Patterson, Garrard, & Hodges, 2000). In contrast, non-verbal reasoning, perceptual and spatial skills, new episodic learning, syntax and phonology remain largely intact (Hodges, Patterson, Oxbury, & Funnell, 1992). SD patients rarely produce phonological errors in spontaneous speech or naming and perform well on phonological tasks like minimal pair discrimination and single word repetition (Jefferies, Jones, Bateman, & Lambon Ralph, submitted-b;

* Corresponding author. Present address: Department of Psychology, University of Manchester, Oxford Road, Manchester M13 9PL, UK.
Tel.: +44-161-275-7338; fax: +44-161-275-2588.
E-mail address: beth.jefferies@man.ac.uk (E. Jefferies).

Knott et al., 1997). Despite intact phonology, ISR for short lists of words is characterised by numerous phoneme migration errors (e.g., "mint, rug" recalled as "rint, mug"), and consequently word span is severely impaired (McCarthy & Warrington, 1987; Patterson, Graham, & Hodges, 1994; Warrington, 1975).

Several studies have found that words are more likely to disintegrate in ISR if their meanings are degraded, supporting the notion that long-term semantic representations play a role in maintaining the phonological coherence of words in STM (Caza, Belleville, & Gilbert, 2002; Forde & Humphreys, 2002; Knott et al., 1997; Knott, Patterson, & Hodges, 2000; Patterson et al., 1994). Patterson et al. (1994) selected words that were either relatively well or poorly understood by individual SD patients, and found that fewer phonological errors occurred during recall of the better-known words. This difference remained when the 'known' and 'degraded' words were matched for frequency on an item-by-item basis (Knott et al., 1997, 2000). Normal participants show similar phoneme migration errors in ISR for non-words, which by definition lack lexical and semantic representations (Treiman & Danis, 1988). It should be noted, however, that a number of studies have failed to find the expected ISR advantage for known over degraded items in semantically impaired patients (Funnell, 1996; Lambon Ralph & Howard, 2000; McCarthy & Warrington, 1987, 2001; Warrington, 1975). Methodological factors contribute to this inconsistency; in particular, the known/degraded difference is less likely to emerge for word lists selected from a small pool of items (Jefferies, Jones, Bateman, & Lambon Ralph, submitted-c).

1.1. The case of number words

Word span is severely impaired in SD but digit span is typically unimpaired. Warrington (1975) described two patients with word spans of four but digit spans of nine and seven. Similarly, patient AM (Knott et al., 1997) recalled lists of five digits almost perfectly but only a quarter of lists of the same length when they were composed of letters or high frequency words. Digit span also remains relatively stable in the face of marked semantic decline (Knott et al., 2000).

This striking difference between digit span and word span could arise for a number of reasons. One intriguing possibility is that it is equivalent to the ISR difference between known and degraded words. That is, ISR for digits might be relatively preserved because they are understood well in comparison with other categories of word. In line with this suggestion, some recent studies have found that number knowledge is relatively spared in patients with SD (Butterworth, Cappelletti, & Kopelman, 2001; Cappelletti, Butterworth, & Kopelman, 2001, 2002; Crutch & Warrington, 2002; Diesfeldt, 1993). The cortical atrophy in SD predominantly affects the anterior and inferior temporal lobes bilaterally, and the temporal poles in particular (Galton et al., 2001; Mummery et al., 2000). In contrast, functional imaging (e.g., Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Dehaene et al., 1996; Stanescu-Cosson et al., 2000) and neuropsychological studies (e.g., Cipolotti, Butterworth, & Denes, 1991; Dehaene & Cohen, 1997; Delazer & Benke, 1997; Warrington, 1982) suggest that knowledge of numbers is associated with the inferior parietal area. Some aspects of number representation may therefore be independent of the temporal lobe semantic system that degrades in SD.

The relative preservation of number knowledge in SD is not the only possible cause of the difference between digit and word span. Single-digit number words are higher in frequency than the words typically used to assess span and ISR performance in SD is strongly affected by frequency (Knott et al., 1997, 2000; McCarthy & Warrington, 2001). In addition, SD patients appear to show enhanced effects of imageability in ISR (Jefferies, Frankish, & Lambon Ralph, submitted-a; Knott et al., 1997) and the words and digits used in span tests are not typically matched for imageability. Thirdly, word span is normally tested with non-repeating items whereas digit span uses a restricted set of nine items. Smaller set sizes improve recall in SD patients as well as normal participants (Jefferies, Jones et al., submitted-c; Knott et al., 1997), so set size may contribute to the better ISR for digits. In addition, digits are drawn from a closed semantic set where as the words used in span tasks can be drawn from many semantic categories. A fifth potential factor is that numbers form an ordered sequence whereas words do not. Finally, normal subjects show better recall of digits than words (Brener, 1940) and it is not clear from the existing reports whether the difference between digit and word ISR in SD patients is more substantial than in controls.

This study examined ISR of number and non-number words matched for frequency, imageability, word length, set size, and size of semantic category in four SD patients, and made a direct comparison between patients and controls. We extended the materials to determine whether better performance with single-digit numbers generalised to lower frequency multi-digit numbers. Comprehension of the number and matched non-number words was assessed in order to examine whether a category-specific difference in comprehension could underlie the ISR results.

2. Case descriptions

The experimental investigations, conducted in 2001, examined four SD patients who are described in order of severity. A summary of the background neuropsychological assessment is shown in Table 1. EK was a 60-year-old right-handed woman who left school at the age of 15 and had been experiencing worsening word-finding difficulties for around 5 years. She was living alone and doing occasional cooking and cleaning jobs at the time of the study. An MRI scan obtained in 2002 showed bilateral temporal lobe atrophy that was more marked in the left hemisphere.

Table 1
Background neuropsychological scores

Test	Maximum	EK	GT	PD	MK	Control mean (S.D.)
Mini-Mental State Examination (Folstein, Folstein & McHugh, 1975)	30	27	26	13[a]	21[a]	Cutoff > 24
Raven's Coloured Progressive Matrices (Raven, 1962)	36	33	35	25[a]	22[a]	–
Digit span: forwards (Wechsler, 1987)	–	6	6	7	5	6.8 (0.9)[b]
Digit span: backwards (Wechsler, 1987)	–	7	4	5	4	4.7 (1.2)[b]
Spatial span: forwards (Wechsler, 1997)	–	6	5	–	5	5–6[c]
Naming	64	17[a]	11[a]	4[a]	2[a]	62.3 (1.6)[b]
Word-picture matching	64	46[a]	32[a]	17[a]	11[a]	63.7 (0.5)[b]
Pyramids and Palm Trees test: pictures (Howard & Patterson, 1992)	52	35[a]	37[a]	26[a]	33[a]	51.1 (1.1)[b]
Pyramids and Palm Trees test: words (Howard & Patterson, 1992)	52	36[a]	32[a]	26[a]	26[a]	51.2 (1.4)[b]
Category fluency: living	–	13[a]	13[a]	5[a]	0[a]	60.3 (12.6)[b]
Category fluency: man-made	–	18[a]	11[a]	2[a]	1[a]	54.8 (10.3)[b]
Letter fluency (F, A, S)	–	29	24	22	2[a]	44.2 (11.2)[b]
Rey figure immediate copy (Lezak, 1976)	36	34	34	36	30	34.0 (3.0)[d]
VOSP[3]: incomplete letters (Warrington & James, 1991)	20	20	18	3[a]	10[a]	19.2 (0.8)[b]
VOSP[3]: dot counting (Warrington & James, 1991)	10	10	10	10	10	9.9 (0.3)[b]
VOSP[3]: position discrimination (Warrington & James, 1991)	20	20	20	16[a]	17[a]	19.8 (0.6)[b]
VOSP[3]: cube analysis (Warrington & James, 1991)	10	10	10	5[a]	6	9.7 (2.5)[b]

Figures refer to the number of correct items.

[a] Denotes abnormal performance.

[b] Control data from Bozeat, Lambon Ralph, Patterson, and Hodges (2002).

[c] Normal range for age matched participants.

[d] Control data from Graham, Patterson, and Hodges (2001).

Her neuropsychological profile was dominated by a moderate impairment of semantic memory. She performed poorly on tests requiring comprehension of words and pictures; for example, word-picture matching and the Pyramids and Palm Trees test (Howard & Patterson, 1992). She was severely anomic in spontaneous speech, word fluency tasks and confrontational picture naming. Her naming errors were predominantly omissions and semantic paraphasias. In common with other SD patients, she produced surface dyslexic errors in reading aloud and surface dysgraphic errors in spelling tasks. In contrast to her marked semantic difficulties, she was well oriented in time and place, had excellent episodic memory for recent events, and had no difficulty in remembering appointments. She performed normally on tests of visual-spatial processing from the Visual Object and Space Perception (VOSP) battery (Warrington & James, 1991), and she was able to produce a good immediate copy of the Rey complex figure (Lezak, 1976). Her non-verbal reasoning on the Coloured Progressive Matrices test (Raven, 1962) was normal. Her speech was fluent and syntactically well formed despite her anomia. She had intact single word phonology and she did not make phonological errors in spontaneous speech or picture naming. She had normal spatial STM as assessed by the Corsi block tapping task, and normal verbal STM as measured by forwards and backwards digit span (Wechsler, 1987). Her word span performance, however, was characterised by frequent phonological errors similar to those described by Patterson et al. (1994).

GT, a 71-year-old right-handed male, left school aged 14 and worked as a builder and a technician in a higher education college. At the time of the study, he had been experiencing a gradual decline in his word finding and comprehension for 5 years. An MRI scan from 2002 showed marked bilateral circumscribed temporal lobe atrophy. His cognitive profile was similar to the description of EK above although his semantic impairments were a little more severe. He was impaired on a range of pictorial and verbal tests of semantic memory. In contrast, he was well oriented in time and space, and had intact visual-spatial skills, non-verbal reasoning abilities and memory for recent events. His speech was fluent and syntactically well formed but characterised by anomia and frequent circumlocutions. His conversation was repetitive. He did not make phonological errors in spontaneous speech or picture naming. He had good verbal STM as measured by forwards and backwards digit span, although his word span performance was characterised by frequent phonological errors. His hearing was slightly impaired in his right ear.

PD, a 73-year-old right-handed woman, left school at the age of 14 and later worked as a regional organiser for a large charity. She had an 8-year history of worsening semantic memory problems and these were very severe at the time of testing. An MRI scan from 1997 showed very marked bilateral temporal lobe atrophy that was worse in the right hemisphere, with relative preservation of more medial temporal-lobe structures including the hippocampus and also evidence of some more generalised cortical atrophy. PD was near floor on a range of tests that required comprehension of pictures and words. Early in the course of the disease, she experienced particular problems with recognising objects and people, and at the time of testing, she showed poorer performance on pictorial compared with verbal semantic tests, consistent with her predominantly right-sided atrophy (Evans, Heggs, Antoun, & Hodges, 1995). Although

she had been well oriented for time and place when she first presented in 1996, she was more poorly oriented at the time of testing and occasionally became lost. She also showed some impairment in visual-spatial skills and non-verbal reasoning. PD exhibited some behavioural changes, including disinhibition, which would be consistent with the disease process affecting basal frontal as well as temporal regions (Snowden, Neary, & Mann, 1996). However, she did not make phonological errors in either picture naming or spontaneous speech. She became increasingly difficult to test and withdrew from the study before all the experiments reported here were completed.

MK, a 67-year-old right-handed woman, was the most severely semantically impaired patient included in the study. She left school at the age of 17 and had previously been employed in clerical work. Her family reported a 3-year history of worsening semantic problems. An MRI scan from 2000 showed marked temporal lobe atrophy that was strongly lateralised to the left side. She performed at or near floor on tests of semantic memory and was severely anomic in both picture naming and spontaneous speech. However, she frequently used numbers in conversation; for example, to describe the age of her grandchildren. In contrast to her semantic impairments, she remained well oriented in time and place, and her memory for recent events was excellent. Her verbal STM was normal as assessed by forwards and backwards digit span. At the time of testing, she appeared to have good single word phonology. She was impaired on tests of non-verbal reasoning and visuospatial processing, but she did not show signs of disinhibition or other behavioural changes.

3. Experiment 1: immediate serial recall of single-digit numbers and matched words

This experiment aimed to determine whether the ISR difference between single-digit numbers and non-number words would persist when the items were matched for length, frequency, imageability and set size.

3.1. Method

Nine words were matched on an item-by-item basis to the digits one to nine for syllable length and word frequency using lemma counts from the Celex database (Baayen, Piepenbrock, & van Rijn, 1993). Many of these high frequency words were abstract in nature, so a second set of words was selected to match for frequency, imageability and syllable length using imageability counts from the MRC psycholinguistic database (Coltheart, 1981). These two sets of words are reproduced in Appendix A.

A variety of list lengths were tested. There were ten lists at each length. The length of the lists depended on each patient's ISR abilities. EK was tested on four to seven items. GT was tested on four to eight items. PD had an exceptional digit span and was tested on six to eight items (although for the second set of words, eight item lists were not tested due to time constraints). MK performed more poorly on ISR tasks, so was tested on four to six items. The numbers and words were yoked so that matched items appeared in the same position in each list. The digits and frequency-matched words were presented in blocks using an ABBA design. The frequency and imageability-matched words were tested on a separate occasion. Testing was extended to include shorter list lengths for words and longer list lengths for digits, allowing the patients' errors just beyond span to be compared across the materials (although this process was not completed for PD). The items in this and subsequent experiments were read aloud at a rate of one word per second for immediate serial recall.

Twelve healthy control participants were also tested on these materials. Three control participants were matched to each of the four patients on age and years of education. They were tested on list lengths from four to nine items for all three types of material. The materials were presented in blocks using an ABCCBA design.

3.2. Results

3.2.1. Recall accuracy

Fig. 1 shows the percentage of single-digit numbers and words recalled in the correct order by patients and controls. Table 2 indicates span for all the materials tested in this series of experiments, and shows that the pattern of results was similar whether performance was measured in terms of list or item accuracy.

The controls recalled the single-digit numbers better than the words, even though they were matched for frequency, imageability and set size. The advantage for numbers was reliable relative to both the first set of frequency matched words ($t(11) = 7.63$, $P < 0.001$) and the second set of frequency and imageability matched words ($t(11) = 9.10$, $P < 0.001$). There was no difference in recall between the two sets of non-number words ($t(11) < 1$). The patients also showed better recall of the numbers than the words. For all four patients, statistical contrasts between the recall of single-digit numbers and both sets of non-number words were significant at $P < 0.001$, with independent-samples t-values ranging from 3.43 to 10.65. None of the patients showed a significant recall difference between the two sets of non-number words (all $t < 1$). These analyses collapsed across list length.

Fig. 1 shows that the patients had normal single-digit recall but impaired word recall relative to the controls. The recall of single-digit numbers was within the normal range for EK, GT and MK, while PD actually outperformed the controls. In contrast, GT and MK had markedly impaired word recall that fell below the normal range. EK's word recall scores were right at the bottom of the normal range. PD's word recall was within the normal range, but she had an exceptional digit span, and the *difference* between digits
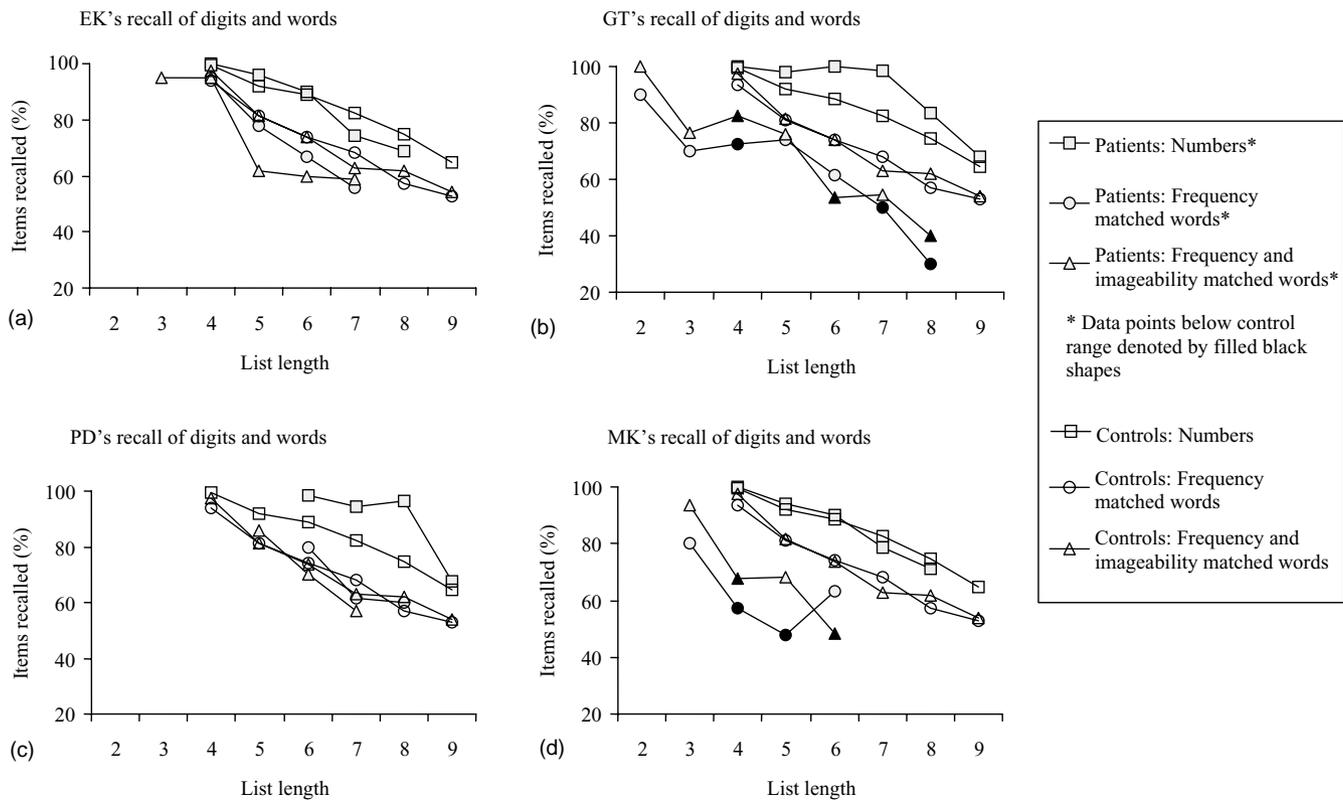
Fig. 1. (a–d) Recall of single-digit numbers and matched words by patients and controls in Experiment 1.

and words fell outside the normal range on some list lengths. The maximum difference between the percentage of digits and words recalled by the control participants was 36%. The largest difference was 34% for EK, 54% for GT, 37% for PD and 46% for MK.

3.2.2. Errors committed on single-digit numbers and words

If the difference between digit and word span corresponds to an ISR difference between known and degraded items, fewer phonological errors should occur in digit span, as the more robust long-term representations of these items should help to hold their phonology in place (Patterson et al., 1994).

There was, however, no straightforward way to compare the patients' digit and word errors because, except for very long list lengths, the patients were at ceiling on digit recall. On lengths where the patients made enough digit errors to analyse, their performance on the word items was abysmal. If digit and word errors were compared at a fixed length, observed differences could result from the discrepancy in difficulty. Therefore, digit and word errors were compared on list lengths that were just above span for the two types of material.

Incorrect responses were categorised as omission, order, repetition, intrusion, phonological and unrelated errors.

Table 2
Spans for the number and non-number words used in Experiments 1–3

	EK	GT	PD	MK	Control mean (range)
Single-digit numbers (Experiment 1)	6	7	8	6	6.08 (5–8)
High frequency words 1 (Experiment 1)	4	2[a]	≤5[b]	3[a]	4.75 (4–7)
High frequency words 2 (Experiment 1)	4	3[a]	5	3[a]	4.67 (4–6)
Low frequency numbers (Experiment 2)	3	3	–	2[a]	3.67 (3–4)
Low frequency words 1 (Experiment 2)	2[a]	2[a]	–	1[a]	4.08 (3–5)
Low frequency words 2 (Experiment 2)	3	3	–	2[a]	4.33 (3–5)
Numbers (Experiment 3)	4	4	–	3	3.83 (3–5)
Face-parts (Experiment 3)	3[a]	3[a]	–	1[a]	4.75 (4–6)

Span was defined as the length at which at least half the lists were repeated correctly. Words 1: frequency matched. Words 2: frequency and imageability matched (Experiment 1) or frequency matched and high imageability (Experiment 2).

[a] Denotes abnormal performance.

[b] PD was above span on six items, but shorter lengths were not tested.

Table 3
Errors on single-digit numbers and words in Experiment 1

Condition	EK			GT			PD			MK			Control mean (range/maximum)		
	Dig	W1	W2	Dig	W1	W2	Dig[a]	W1[a]	W2	Dig	W1	W2	Dig	W1	W2
Span	6	4	4	7	2[b]	3[b]	8	≤5	5	6	3[b]	3[b]	6.1 (5–8)	4.8 (4–7)	4.7 (4–6)
No. of items	210	150	150	240	90	120	170	110	180	210	120	120	212.5 (180–270)	172.5 (150–240)	170.0 (150–210)
Phonological	0	5	6[b]	0	18[b]	21[b]	0	21[b]	7[b]	0	28[b]	22[b]	0	2.9 (13)	0.8 (4)
Unrelated	0	1	1	0	2	2	0	0	5	0	9[b]	8[b]	0	0.3 (2)	0.9 (5)
Omission	9	13	14	14	1	0	7.5	18	26[b]	18	5	2	11.3 (35)	13.3 (40)	13.4 (25)
Order	27	12	20	15	0	2	25.5	6	20	13	0	0	14.8 (29)	10.9 (25)	10.3 (20)
Repetition	10	0	2	13	0	1	22.5[b]	10.5	6	6	3	1	7.6 (20)	5.8 (13)	4.7 (16)
Intrusion	3	4	2	2	1	1	0	6	1	8	4	0	3.8 (10)	8.4 (18)	10.2 (22)

Dig: single-digit numbers; W1: frequency matched words; W2: frequency and imageability matched words. Figures refer to the total number of errors.

[a] Fewer data contributed to PD's scores, so they were scaled up (multiplied by 1.5) to allow a rough comparison with the other patients and controls.

[b] Denotes abnormal performance.

Omission errors were calculated by subtracting correct responses and other error types from the number of items presented. Responses were counted as an order error if the word produced was a target word occurring somewhere else in the sequence. Repetition errors were target words produced more than once. Intrusion errors were items presented in a previous list. A phonological error reproduced at least 50% of the phonemes from the target word (e.g., 'bread' → 'bed', 'sorry' → 'forry'). Unrelated errors did not fall into any of these categories. Errors of this type were most commonly patient responses that did not reach the criteria for a phonological error (e.g., 'council' → 'cathert').

Table 3 indicates the total number of errors of each type, for each individual patient and for the controls as a group, across three list lengths: span, and one and two items beyond span. Span was defined as the longest length at which at least 5/10 lists were repeated correctly. PD was not tested on all the lengths necessary for this method. She was not tested on numbers at a length of two items beyond span, or on lengths short enough to obtain span for the first set of words. PD's digit scores are an amalgamation of span and one length beyond span, and her scores on the first set of words are a combination of the two shortest lengths tested. Fewer data contributed to PD's scores, so they were scaled up (multiplied by 1.5) to allow a rough comparison to be made with the other patients.

The error rates of all four patients on were either within the normal range or nearly normal for the recall of single-digit numbers. The majority of errors were omissions, order errors, repetitions and intrusions, for both patients and controls. There was virtually a complete absence of phonological and unrelated errors for both groups. The controls' errors on non-number words followed a similar pattern, although there were more intrusions (word set 1: $t(11) = 3.55$, $P < 0.01$; word set 2: $t(11) = 4.09$, $P < 0.01$) and more unrelated errors (word set 1: $t(11) = 2.91$, $P < 0.05$; word set 2: $t(11) = 2.42$, $P < 0.05$) in word recall.

The four patients showed a pattern of errors in word recall that was very different both from controls' word recall and from their own digit recall. For GT, PD and MK, the number of phonological errors exceeded the control range on both word sets. For EK, the number of phonological errors was outside the control range on the second word set. MK also made a large number of unrelated errors that failed to reach the criterion for a phonological error but may have occurred for similar reasons. In contrast, the numbers of omission, order, repetition and intrusion errors were within the normal range.

For all four patients, there were reliable differences between the pattern of errors in digit and word recall. Out of eight possible contrasts between a patient's pattern of errors in digit versus word recall (digits versus word set 1 and digits versus word set 2 for each of the four patients), all eight revealed a statistically reliable difference at $P < 0.01$ or less, with chi-squared values ranging from 14.5 to 74.6. Furthermore, the standardised residuals for phonological errors on the non-number words were high in all cases, suggesting that this error category was a major contributor to the significant chi-squared values.

3.3. Discussion

The four SD patients had normal digit spans but impaired word spans, even when the materials were matched for frequency, imageability, word length and set size. In addition, phonological errors occurred frequently in word span but extremely rarely in digit span. These differences are reminiscent of those reported previously for known and degraded words (Knott et al., 1997; Patterson et al., 1994).

The results of this experiment leave us with a puzzle—why are SD patients normal at repeating sequences of single-digit number words but impaired at repeating non-number words if this difference cannot be accounted for by frequency, imageability, word length or the number of items in the set? The following experiments investigated other possible reasons for the difference, in particular the idea that SD patients understand number words better than they do non-number words.

4. Experiment 2: immediate serial recall of low frequency multi-digit numbers and matched words

This experiment examined the ISR for lower frequency multi-digit number words, like billion and ninety, together with matched non-number words. Multi-digit numbers are expected to be recalled and comprehended more poorly than single-digit numbers because (1) low frequency words and concepts typically degrade earlier in the course of SD (Funnell, 1995) and (2) multi-digit numbers refer to more difficult numerical concepts (Dehaene & Mehler, 1992). However, lower frequency non-number words should also engender poor recall if semantics makes a major contribution to phonological coherence, and hence the difference between the material types might remain.

This experiment also addressed one concern about the interpretation of the previous study. In Experiment 1, both the healthy controls and the SD patients showed better recall of single-digit numbers than non-number words; therefore, it is possible that the patients' specific ISR impairment for non-number words occurred simply because this task was harder. The use of lower frequency multi-digit numbers and matched words circumvented this problem, because the normal recall advantage for number words was eliminated.

4.1. Method

The nine lowest frequency number words in English that were whole words rather than compounds of words (e.g.,

13, not 37) were selected using the Celex database (Baayen et al., 1993). The numbers were compared with words matched on an item-by-item basis for syllable length and frequency. Imageability ratings were not available for many of these words (using the MRC corpus, Coltheart, 1981), so a second set of frequency matched, high imageability words was selected. The items are reproduced in Appendix B.

As in Experiment 1, the lists of numbers and words were yoked so that matched items appeared in the same position within a list. There were 10 lists at each length. EK and GT were tested on list lengths from two to seven items, and MK was tested on two to six items (as her performance was poorer). In every case, the numbers and both sets of words were tested at each list length. MK was additionally tested on a single item from the first set of non-number words, to determine her span in this condition. PD was not available to participate in this or subsequent experiments. The numbers and frequency matched words were tested in blocks using an ABBA design. The frequency matched, high imageability words were tested separately. The 12 control participants described for Experiment 1 were also tested on these materials, using list lengths from three to eight items arranged in an ABCCBA design.

4.2. Results

4.2.1. Recall accuracy

Fig. 2 indicates the number of items recalled in the correct order by patients and controls. In contrast to Experiment 1,
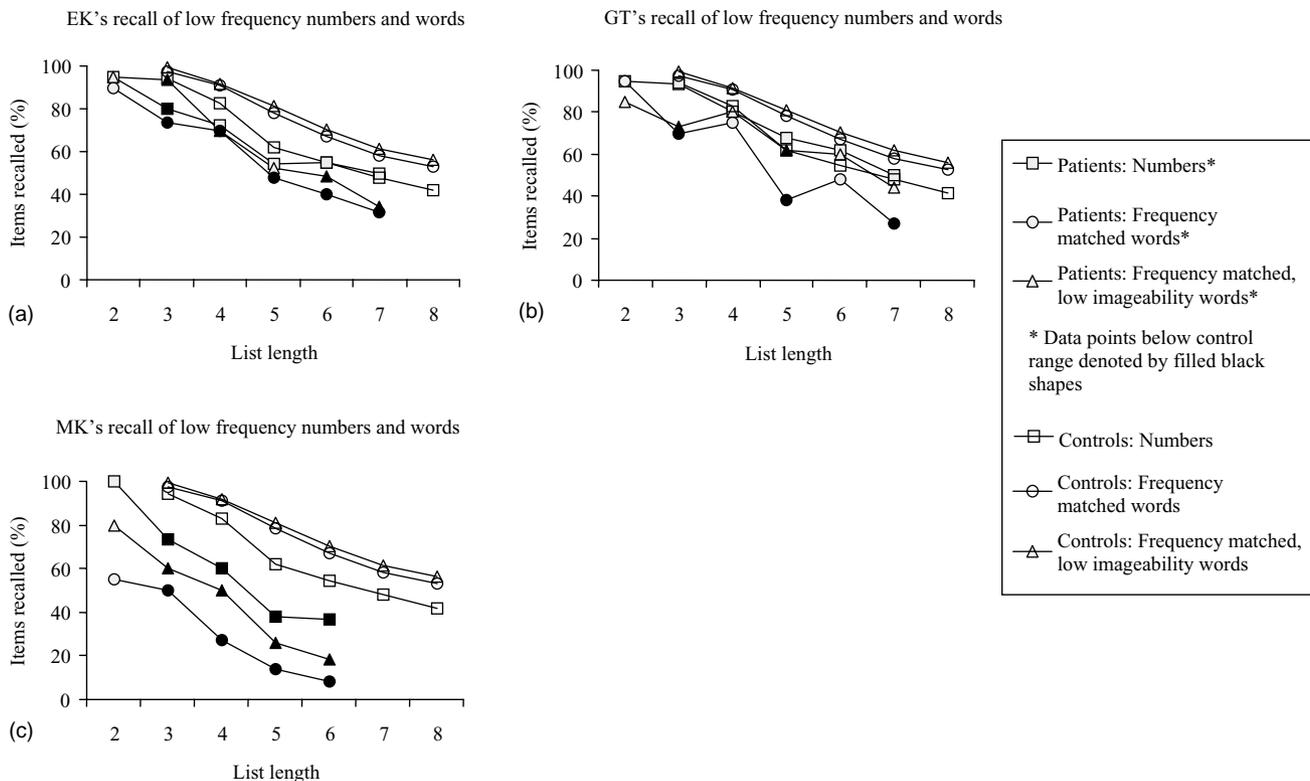

Fig. 2. (a–c) Recall of low frequency multi-digit numbers and matched words by patients and controls in Experiment 2.

the control participants did not show an advantage for repeating multi-digit number words over non-number words. In fact, they showed a highly significant advantage for repeating the non-number words over the low frequency numbers (collapsing across list length: numbers versus word set 1: $t(11) = 6.55$, $P < 0.001$; numbers versus word set 2: $t(11) = 8.74$, $P < 0.001$). In addition, the controls recalled the higher imageability words (set 2) more accurately than the lower imageability words (set 1: $t(11) = 5.51$, $P < 0.001$).

EK showed an ISR advantage for the numbers over the first set of words which approached significance (collapsing across list length: $t(114) = 1.94$, $P < 0.06$). She showed no difference between the numbers and the second set of words ($t(113) < 1$). GT's ISR performance on the numbers was significantly better than for the first set of words ($t(108) = 3.17$, $P < 0.01$) and no different from the second set of words ($t(112) = 1.53$, n.s.). MK recalled the numbers better than the words from both set 1 ($t(98) = 5.07$, $P < 0.0001$) and set 2 ($t(98) = 2.22$, $P < 0.05$). Therefore, all three patients recalled the numbers as well as or better than the non-number words, whereas the controls recalled the non-number words more accurately than the numbers.

ISR for the numbers was within the normal range for GT, whereas EK was mildly impaired (her recall was lower than the control mean, and dipped below the lowest control score on a few list lengths). MK was more substantially below the normal range for number words. However, all three patients showed a much greater impairment in their recall of non-number words. EK and MK were very severely impaired on both sets of non-number words, with recall falling substantially below the normal range on every length tested. Recall of the non-number words was below the normal range for GT in parts of the dataset. Moreover, all three patients showed an advantage for number over word recall that was much larger than the maximum observed in the controls (3%). The maximum advantage for number over word recall was 19% for EK, 30% for GT and 33% for MK.

MK recalled the high imageability words (set 2) more accurately than the low imageability words (set 1) ($t(97) =$

2.63, $P < 0.01$). Neither EK nor GT showed a significant difference between the two sets of words (EK: $t(118) = 1.38$, n.s.; GT: $t(117) = 1.57$, n.s.). However, in all three patients, the difference between the two sets of words was larger on some lengths than the maximum observed in control participants (14%), suggesting an enhanced effect of imageability in the ISR of these patients. This result is consistent with previous findings (Jefferies, Frankish et al., submitted-a; Knott et al., 1997).

4.2.2. Errors committed on numbers and words

Errors were analysed in the same way as for Experiment 1 but with one difference. For the low frequency number lists, both the patients and controls sometimes produced number words that were not in the set, but shared 50% of their phonemes with a target number. These errors met the criterion for a phonological error because of the phonological overlap between numbers like 'thirteen' (in the set) and 'sixteen' (not in the set). However, they did not appear to result from the migration, substitution, addition or deletion of phonemes, and were therefore placed in a separate category of number intrusions from outside the set.

Table 4 gives the number of errors made by the patients and controls in each category, combining across three list lengths: span, span +1 and span +2 items. The controls' recall of the number words was characterised by frequent intrusions, from both inside and outside the set, and by omission and order errors. They made a similar pattern of errors on the low frequency words, although within-set intrusions were less numerous than for numbers (word set 1: $t(11) = 5.67$, $P < 0.001$; word set 2: $t(11) = 2.61$, $P < 0.05$). In addition, there were more omission errors ($t(11) = 3.11$, $P < 0.05$), phonological errors ($t(11) = 2.22$, $P < 0.05$) and unrelated errors ($t(11) = 2.42$, $P < 0.05$) on the lower imageability words (set 1) compared with the numbers.

The patients' errors on the non-number words were different in nature to those made by controls. The number of phonological errors greatly exceeded the normal range on both sets of non-number words for all three patients. In addition, the patients with more severe semantic impairments

Table 4
Errors on low frequency numbers and words for patients and controls in Experiment 2

Condition	EK			GT			MK			Control mean (maximum)		
	Num	W1	W2	Num	W1	W2	Num	W1	W2	Num	W1	W2
Span	3	2[a]	3	3	2[a]	3	2[a]	1[a]	2[a]	3.67 (3–4)	4.1 (3–5)	4.3 (3–5)
No. of items	120	90	120	120	90	120	90	60	90	140 (120–150)	152.5 (120–180)	160 (120–180)
Phonological	0	19[a]	8[a]	4[a]	17[a]	26[a]	4[a]	22[a]	24[a]	0.2 (2)	2.8 (12)	0.4 (2)
Unrelated	0	0	1	0	1	5[a]	1	2	3[a]	0.2 (1)	1.0 (3)	0.1 (1)
Omission	7	0	15	6	2	1	4	0	8	9.3 (17)	15.2 (29)	13.9 (30)
Order	6	1	9	1	0	1	0	0	0	8.8 (14)	8.8 (19)	10.0 (17)
Repetition	2	0	1	0	0	1	0	0	1	3.7 (8)	3.6 (10)	3.8 (10)
Within set intrusion	12	0	2	10	0	2	13	0	0	12.8 (19)	6.6 (17)	9.1 (21)
Number intrusion	12	–	–	5	–	–	5	–	–	10.3 (17)	–	–

Num: numbers; W1: frequency matched words; W2: frequency matched, high imageability words. Figures refer to the total number of errors.

[a] Denotes abnormal performance.

made larger numbers of phonological errors. In contrast, the numbers of omission, order, repetition, and intrusion errors did not exceed the normal range. The number of unrelated errors also exceeded the normal range for GT and MK in the second set of non-number words. The patients' errors on the multi-digit number words were more similar to those made by controls. The numbers of omission, order, repetition and intrusion errors (from within and outside the set) did not exceed the normal range. However, the more severely impaired patients, GT and MK, made slightly more phonological errors on the number words than the controls. Evidence is presented below to suggest the patients' comprehension of low frequency multi-digit numbers was impaired, consistent with an association between semantics and phonological errors in ISR within the number domain.

As in Experiment 1, there were reliable differences between the pattern of errors in number and word recall for every patient. Out of six possible contrasts between the patients' pattern of errors in number versus word recall (numbers versus word set 1 and numbers versus word set 2 for each of the three patients), all six revealed a statistically reliable difference at $P < 0.001$, with chi-squared values ranging from 27.3 to 57.1. The standardised residuals for phonological errors on the non-number words were high in every case.

4.3. Discussion

Two patients were mildly impaired at recalling the low frequency multi-digit numbers relative to controls, but all three patients were much more impaired at recalling the matched non-number words, making the difference between the materials greater in the patients than controls. Moreover, the quantity and quality of the patients' errors were similar to those of the control participants on the number words, but the patients made many more phonological errors than the controls on the non-number words. The patients' abnormal ISR advantage for numbers extended beyond single digits to low frequency multi-digit numbers. This is a potentially important finding because digit span can be preserved in patients with otherwise severe aphasia (Cohen, Verstichel, & Dehaene, 1997), suggesting that the ability to repeat digits may be over-learned or automated and therefore protected. It seems unlikely that low frequency numbers could be automated in the same way and therefore this possibility does not provide an adequate account of the data.

5. Experiment 3: immediate serial recall of numbers and face-part words

A third experiment examined recall of middle frequency, mostly multi-digit numbers and frequency-matched words that loosely fitted into the category of 'face or head parts' (for example, mouth, fringe, beard). This study had several aims. First, comprehension of the number and face-part words

could be directly compared in naming and word-picture matching tasks, allowing us to investigate whether the ISR difference between number and non-number words corresponded to a difference in comprehension. The results of these semantic tasks are discussed in a separate section below. Secondly, both the number and face-part words were drawn from closed semantic categories. In the experiments above, the number words were drawn from a single semantic category, whereas the non-number words were drawn from many semantic categories, making the non-number words less predictable. This experiment examined whether the superior recall of number words would persist after matching for this feature. Thirdly, it could be argued that the numbers in Experiment 1 were more imageable than the words they were matched with, because it is apparently easier to form a mental image of an Arabic numeral, e.g., '3', than a word with an intermediate imageability rating, e.g., 'small'. However, the face part words used in this experiment were, according to published ratings, considerably more imageable than the digits 1 to 9. Consequently, if the advantage for number words persists in this experiment, it is unlikely to result from enhanced imageability effects in the patient group.

5.1. Method

Twelve number words (whole words rather than compounds) were compared with twelve face-part words in an ISR task. The items were matched as closely as possible for frequency using the Celex database (Baayen et al., 1993). It did not prove possible to match the groups for syllable length; the number words were significantly longer. The face-part words also had higher imageability ratings in the MRC online corpus (Coltheart, 1981; see above). However, these two differences should reduce the recall advantage for number words shown by SD patients. The items are reproduced in Appendix C.

EK was tested on lists containing three to seven items. GT and MK were tested on lists containing two to seven items. MK was additionally tested on a single face part word. The twelve controls were tested on lists containing three to eight items. Ten lists were tested at each length. In list construction, the number and face-part words were yoked so that matched items appeared in the same position in the lists. The numbers and face-part words were tested in blocks using an ABBA design.

5.2. Results

5.2.1. Recall accuracy

Fig. 3 shows the percentage of number and face-part items recalled in the correct order by the patients and controls, and Table 2 indicates span for these materials. The controls recalled the face-part words better than the numbers, consistent with the shorter length and higher imageability of the
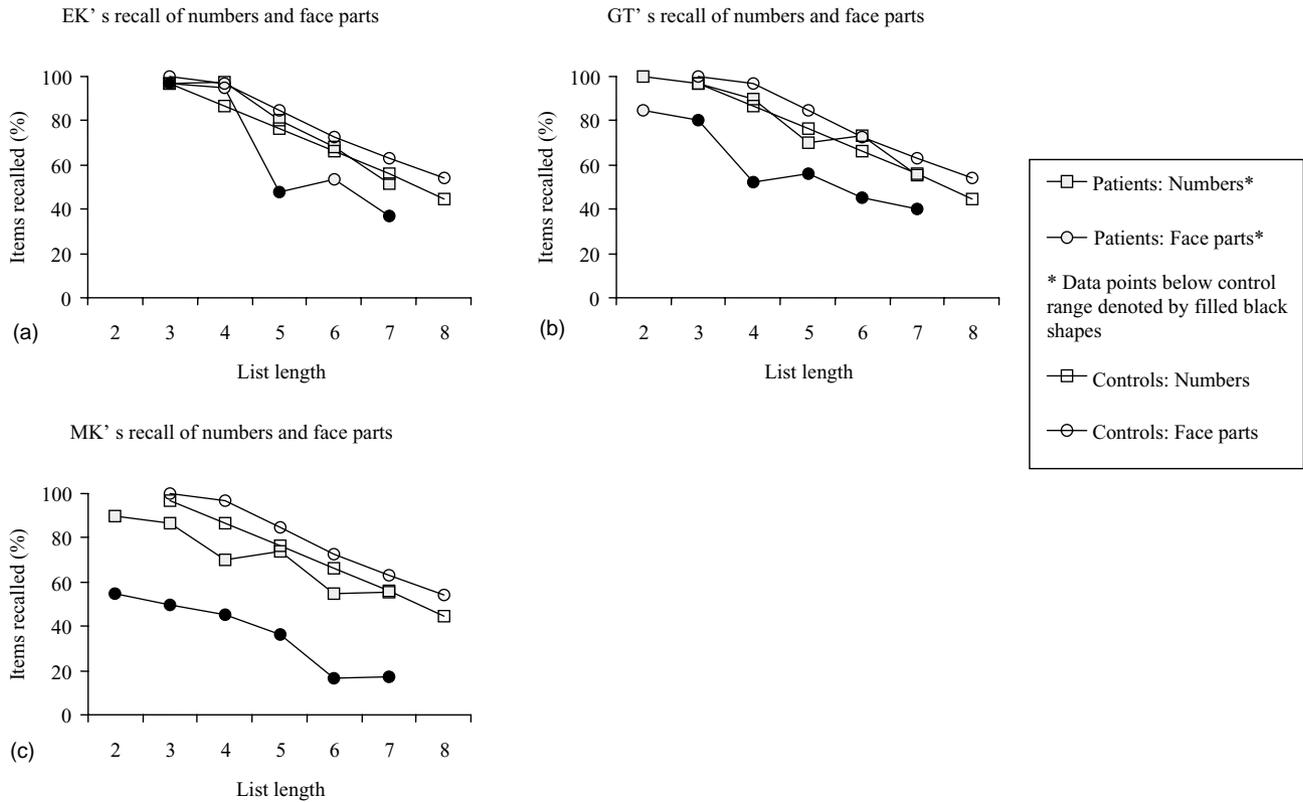
EK' s recall of numbers and face parts

GT' s recall of numbers and face parts

MK' s recall of numbers and face parts

Fig. 3. (a–c) Recall of face parts and matched number words by patients and controls in Experiment 3.

items in the former set ($t(11) = 7.94$, $P < 0.001$). In contrast, the patients showed better recall of the numbers than the face-parts (EK: $t(92) = 3.24$, $P < 0.01$; GT: $t(114) = 4.51$, $P < 0.0001$; MK: $t(116) = 7.40$, $P < 0.0001$). These analyses collapsed across list length. Recall of the number items was within the normal range for all three patients on every length tested. In contrast, recall of the face-part words was below the normal range for all three patients, on almost every length tested. Moreover, all three patients showed an advantage for number over word recall that was much larger than the maximum observed in the controls (32% for EK, 38% for GT and MK, 5% for the controls).

5.2.2. Errors on number and face-part words

Errors were categorised as for Experiment 1, with the additional category of 'outside-set intrusions' (i.e., number words and parts of the head or face that were not included in the experimental set). Errors were placed in these categories even if they also met the criteria for a phonological error. Table 5 shows the number of errors of each type produced by patients and controls, combining across span, span +1 and span +2 list lengths.

The controls largely made intrusion, omission and order errors for both numbers and face-parts. Several error types were more numerous for face-parts than for numbers,

Table 5
Errors on numbers and face-part words for patients and controls in Experiment 3

Condition	EK		GT		MK		Control mean (max)	
	Num	Word	Num	Word	Num	Word	Num	Word
Span	4	3[a]	4	3[a]	3	1[a]	3.8 (3–5)	4.75 (4–6)
No. of items	150	120	150	120	120	60	145 (120–180)	172.5 (150–210)
Phonological	1	17[a]	1	34[a]	0	15[a]	0	0.1 (1)
Unrelated	0	1	0	12[a]	0	9[a]	0	0.1 (1)
Omission	9	15	4	0	4	0	8.3 (19)	15.8 (44)
Order	3	5	5	0	4	0	4.1 (9)	9.8 (18)
Repetition	2	0	3	0	1	0	1.6 (5)	3.6 (9)
Within set intrusion	13	1	14	0	7	0	9.3 (21)	14.1 (25)
Outside set intrusion	10	0	8	0	14	0	10.5 (17)	0.1 (1)

[a] Denotes abnormal performance. Figures refer to the total number of errors.

including omissions ($t(11) = 2.38$, $P < 0.05$), order errors ($t(11) = 3.53$, $P < 0.01$), repetitions ($t(11) = 3.02$, $P < 0.05$) and within-set intrusions ($t(11) = 2.94$, $P < 0.05$). Intrusions from outside the set were more numerous for the numbers ($t(11) = 10.19$, $P < 0.001$).

The three patients had a virtually normal pattern of errors for the number words. Omission, order, repetition, within and outside-set intrusion and unrelated errors were within the normal range. The patients' errors on the face-part words were different in nature. All three patients made a large number of phonological errors that greatly exceeded the normal range, and the number of unrelated errors was also above the normal range for GT and MK. In contrast, the numbers of omission, order, repetition and intrusion errors were within the normal range.

Error patterns on the number and face-part words were significantly different for all three patients (EK: $\chi^2(1) = 42.51$, $P < 0.001$; GT: $\chi^2(1) = 85.79$, $P < 0.001$; MK: $\chi^2(1) = 58.85$, $P < 0.001$). The standardised residuals were high for phonological errors on the face-part words in every case, suggesting that the large number of phonological errors on these words underpinned the difference in error patterns.

5.3. Discussion

As in the previous two experiments, the patients recalled the number words at relatively normal levels, but were significantly impaired at recalling the non-number face-part words. In addition, the patients made many more phonological errors than the controls on the non-number words but not on the number words. The ISR difference between number and non-number words remained in this experiment when both sets of items were drawn from closed semantic categories.

6. Experiment 4: immediate serial recall of letters

Experiments 1–3 suggest that number repetition is relatively intact in SD compared with word repetition, even when the materials are broadly equated for word length, frequency, imageability, set size, and open or closed semantic categories. However, numbers occur in a sequence and are represented by single characters, and these characteristics could also aid recall. Moreover, in some neurological patients, knowledge of number sequence patterns with other ordered series, such as days of the week or months of the year (Cipolotti et al., 1991; Dehaene & Cohen, 1997; Thioux et al., 1998). A fourth experiment compared ISR for letters with the single-digit numbers and matched non-number words used in Experiment 1, as letters share these unusual properties. This experiment also compared phonologically similar and dissimilar letters. If SD patients have intact phonological STM capacities, as is generally assumed, then they, like normal participants, should show poorer recall of phonologically similar than dissimilar items; an effect which is typically attributed to phonological coding in STM (e.g., Conrad & Hull, 1964).

6.1. Method

Following Knott et al. (2000), the patients and twelve controls were asked to repeat lists of four and six letters from the phonologically similar set E, C, T, P, V, B, G, D or the phonologically dissimilar set S, Q, Y, R, J, F, W, L. The set size was limited to eight items because there are only eight phonologically similar letters. There were twenty lists at each length divided equally between the phonologically similar and dissimilar sets. The similar and dissimilar letters were presented in blocks using an ABBA design.

6.2. Results

Table 6 shows the number of phonologically similar and dissimilar letters recalled in the correct order, together with scores for single-digit number words and high frequency non-number words (second set) from Experiment 1.

6.2.1. Comparison of letter, digit and word recall

The controls' letter recall was intermediate between ISR for single-digit numbers and non-number words. They recalled the digits better than both the phonologically similar and dissimilar letters ($t(11) = 8.37$, $P < 0.0001$ and $t(11) = 8.37$, $P < 0.0001$, respectively). The high frequency words

Table 6
ISR for letters, digits and frequency-matched words

	EK	GT	MK	Controls mean (range)
Four similar letters	90	85	73	82.7 (63–98)
Four dissimilar letters	100	90	83[a]	96.5 (85–100)
Four digits	100	100	100	99.6 (95–100)
Four words (set 2, Experiment 1)	95	83[a]	68[a]	97.5 (87–100)
Six similar letters	68	75	53	70.1 (53–82)
Six dissimilar letters	87	82	72	83.8 (72–97)
Six dissimilar–similar letters	18	7	18	13.6 (5–23)
Six digits	90	100	90	88.7 (70–100)
Six words (set 2, Experiment 1)	60	53[a]	48[a]	74.0 (55–90)

Note: Figures indicate percentage of items recalled in correct order.
[a] Denotes abnormal performance.

were recalled more poorly than the phonologically dissimilar letters ($t(11) = 3.53$, $P < 0.01$) but more accurately than the similar letters ($t(11) = 4.15$, $P < 0.01$).

The patients showed a comparable pattern. All three patients recalled the single-digit numbers at a higher level than the phonologically similar letters (EK: $t(32) = 3.71$, $P < 0.001$; GT: $t(19) = 5.90$, $P < 0.0001$; MK: $t(35) = 7.84$, $P < 0.001$). Two of the patients also showed better ISR for digits than phonologically dissimilar letters (GT: $t(19) = 3.90$, $P < 0.001$; MK: $t(33) = 4.02$, $P < 0.001$; EK: $t(38) < 1$). In addition, in every case, the phonologically dissimilar letters were recalled more accurately than the non-number words (EK: $t(27) = 2.80$, $P < 0.01$; GT: $t(35) = 2.96$, $P < 0.01$; MK: $t(36) = 3.31$, $P < 0.01$). GT showed better recall of phonologically similar letters than non-number words ($t(34) = 2.04$, $P < 0.05$), whereas EK and MK showed no difference between these conditions (EK: $t(32) < 1$; MK: $t(34) = 1.01$, n.s.). These analyses combined data from both list lengths.

Table 6 indicates that EK and GT showed normal recall of phonologically similar and dissimilar letters. Therefore, these two patients had intact digit and letter repetition abilities, but impaired word repetition abilities. The most severely impaired patient, MK, had mildly impaired letter repetition abilities in the context of severely impaired word repetition and normal digit repetition. MK, unlike the other two patients, was impaired on the degraded letters sub-test in the VOSP. She was only able to name ten out of twenty degraded letters, suggesting that she may have had difficulty recognising their visual forms. MK was also unique in that she made a number of phonological errors and non-letter intrusions in the letter span task (for example, she recalled G as 'chee', and she recalled Q as 'car' and then changed it to 'R'). Therefore, it seems that MK's letter repetition may have been impaired because her knowledge of letters was degraded.

6.2.2. Phonological similarity effects

The controls showed a highly significant effect of phonological similarity ($t(11) = 7.94$, $P < 0.0001$), combining across list lengths. EK and MK also showed better recall of phonologically dissimilar items ($t(35) = 3.33$, $P < 0.01$ and $t(37) = 2.86$, $P < 0.01$ respectively). The numerical difference between phonologically similar and dissimilar items failed to reach significance for GT ($t(38) = 1.17$, n.s.). However, the recall difference between phonologically similar and dissimilar letters was within the normal range for all three patients (see Table 6).

6.3. Discussion

The patients' relatively intact recall of letters is consistent with the notion that ISR is selectively preserved for items that form an ordered series. It is possible that ordered sequences are preserved in SD because they derive support

from a intact understanding of spatial relations (see Gevers, Reynvoet, & Fias, 2003).

All three patients exhibited an effect of phonological similarity that fell within the normal range, consistent with the suggestion that SD patients use normal phonological encoding in STM and that semantic rather than phonological impairments are the cause of their poor verbal STM (see Jefferies, Jones et al., submitted-b). This finding also suggests that any differences in phonological similarity between the number and non-number words used in Experiments 1–3 should have had comparable effects on the ISR of the patients and controls.

7. Number comprehension and processing

Experiments 1–4 indicate that SD patients have relatively intact ISR for numbers and more impaired ISR for non-number words. Substantial and reliable ISR differences were observed for every patient, and these were largest in the patient with the most severe semantic impairment. The ISR difference was not eliminated by matching for frequency, imageability, word length, set size or open versus closed semantic category, suggesting that these factors cannot account for the pattern of results. This section examines the patients' knowledge of numbers to investigate the possibility that ISR for these items is specifically preserved because comprehension of numbers is relatively intact in SD.

There were four elements to these investigations of number processing. First, naming and word-picture matching tasks were devised for the number and face-part words used in Experiment 3, to establish if the superior ISR for numbers corresponded to better comprehension. Secondly, the patients' ability to transcode between Arabic numbers and spoken number words was assessed for the items used in Experiments 1–3. Thirdly, the patients' understanding of the numbers used in the first three experiments was assessed using sequence and magnitude judgement tasks, providing a means of evaluating the idea that their poorer recall of lower frequency multi-digit numbers corresponded with poorer comprehension. Finally, the patients' abilities to understand and manipulate numbers were explored more generally. On most of these tests, the performance of controls was at ceiling, and therefore only one age and education-matched control was tested for each patient unless otherwise stated.

7.1. Comprehension of number and face-part words

7.1.1. Method

Naming and word-picture matching tests were devised for the number and face-part words used in Experiment 3. The numbers were represented pictorially as dots; tens were depicted as clusters of ten dots, and units were depicted as single dots. The patients were asked to provide a name for each dot picture in turn. They were also shown the complete set of twelve pictures together and were asked to point to

Table 7
Naming and word-picture matching using the number and face-part words from Experiment 3

	Maximum	EK	GT	MK
Naming: numbers	12	12	12	12
Naming: face-parts	11	6[a]	6[a]	3[a]
Word-picture matching: numbers	12	12	12	11[a]
Word-picture matching: face-parts	11	9[a]	10[a]	5[a]

[a] Denotes abnormal performance.

the picture that represented a particular number word. The face-part tests used a picture of an entire face. The patients were asked to name the face-parts that were indicated by arrows. In word-picture matching, the patients were asked to select the arrow that pointed to a particular face-part. The item 'brow' was replaced by another face-part in the word-picture matching test, as it was not pictorially distinct from 'forehead'.

7.1.2. Results

The three controls performed without error. Table 7 gives the results for the patients. All three patients were virtually at ceiling on naming and word-picture matching with number words. In contrast, all the patients were impaired on the face-parts. Their naming errors consisted of omissions and semantically-related responses. The patients performed significantly better with the numbers than with the face-parts when the results for naming and word-picture matching were combined (EK, $\chi^2(1) = 6.71$, $P < 0.01$; GT, $\chi^2(1) = 5.32$, $P < 0.05$; MK, $\chi^2(1) = 15.87$, $P < 0.0001$).

7.1.3. Discussion

There was a comprehension difference between the number and face-part words, even though these items were matched for frequency. This suggests that the number do-

main may be relatively intact in SD and that this is the primary basis for the ISR difference between number and non-number words.

7.2. Transcoding of single-digit and lower frequency multi-digit numbers

7.2.1. Method

The patients were asked to read aloud and write Arabic numerals for each of the number words used in Experiments 1–3. An inability to perform these transcoding tasks might indicate that the patients did not comprehend the words used in the ISR experiments. There is controversy, however, about the extent to which transcoding tasks rely on 'semantic' representations of number (Cipolotti & Butterworth, 1995; Dehaene, 1992; Dehaene & Cohen, 1995; Deloche & Seron, 1982).

7.2.2. Results

The results are summarised in Table 8. The patients were able to translate between spoken number words and Arabic numbers almost without error; only the items 'billion' and 'trillion' created difficulties. The milder patients showed some understanding that these words represented large numbers (e.g., EK wrote billion as 10000).

7.3. Comprehension of single-digit and lower frequency multi-digit numbers

7.3.1. Method

Two tests examined comprehension of the single-digit number words used in Experiment 1 relative to the lower frequency multi-digit numbers used in Experiments 2 and 3. First, the patients were asked to arrange cards with the number words printed on them in numerical order. The number

Table 8
Comprehension and processing of the number words used in Experiments 1–3

		Experiment 1: single-digit	Experiment 2: low frequency multi-digit	Experiment 3: medium frequency, mostly multi-digit
	Maximum	9	9	12
EK	Reading numerals	9	7	12
	Writing numerals	9	7	12
	Ordering numbers	9	6[a]	11[a]
	Which number is closest?	8[a]	7[a]	8[a]
GT	Reading numerals	9	7	12
	Writing numerals	9	7	12
	Ordering numbers	9	8[a]	12
	Which number is closest?	9	7[a]	11[a]
MK	Reading numerals	9	7	12
	Writing numerals	9	7	12
	Ordering numbers	9	3[a]	10[a]
	Which number is closest?	7[a]	2[a]	8[a]
Controls	Reading numerals	9	9	12
	Writing numerals	9	7–9	12
	Ordering numbers	9	9	12
	Which number is closest?	9	9	12

[a] Denotes abnormal performance.

words were read aloud by the experimenter throughout the test. Secondly, the patients were asked to select the number out of four that was closest to a target. For example, they were asked 'Which number is nearest to five: eight, nine, two or three?'. The correct response could be either larger or smaller than the target. The alternatives were drawn from the same experimental set of numbers as the target. The numbers were simultaneously read aloud and presented as written number words. It should be noted that these tasks may tap separable aspects of number knowledge (Delazer & Butterworth, 1997).

7.3.2. Results

The results are summarised in Table 8. The patients were able to place the digit words in the correct order, but unlike controls, they made some errors on the mid-frequency numbers from Experiment 3, and a larger number of errors on the low frequency numbers from Experiment 2. Some of the patients' errors appeared to result from confusions between '-teen' words like 'thirteen' and 'fourteen', and '-ty' words like 'thirty' and 'forty' (e.g., $16 \rightarrow 70 \rightarrow 18$) but some did not (e.g., $80 \rightarrow 90 \rightarrow 70 \rightarrow 19$). MK was very slow at this task and appeared to be using an ineffective 'counting up' strategy that was successful with single-digit numbers but not larger numbers.

The patients performed even more poorly on the 'which number is nearest' task. EK and MK made errors on this task even when it involved single-digit numbers. When the patients made errors, they largely selected the distracter second nearest to the target for single-digit numbers and medium frequency numbers (9/12 errors across the patients), suggesting that they knew, at least approximately, about the sequence of numbers and their magnitudes. They showed this pattern less often for low frequency multi-digit numbers (2/11 errors across the patients).

7.3.3. Discussion

There is evidence for an association between the degree of semantic degradation and ISR within the numbers domain itself: the patients' understanding of lower frequency numbers was poorer than their understanding of single-digit numbers. Similarly, they showed poorer recall of these items and made more frequent phonological errors on them. The patients performed at ceiling on one task involving single-digit numbers but they made a larger number of errors in a second task that involved calculation. This finding is consistent with the view that an understanding of numerical magnitude is largely intact in SD but that knowledge of calculation procedures is impaired (see Cappelletti et al., 2001). Some additional assessments that provide some support for this view are presented below.

7.4. Other tests of numerical abilities

We collected some additional data about the patients' numerical abilities that did not relate specifically to the sets of numbers used in Experiments 1–3. First, the patients' number knowledge was examined using a numerical comparison task. In addition, their calculation abilities were assessed in two tasks; they were given arithmetic questions to solve (e.g., '$11 + 8$', '4×1') and they were asked to provide the next number in sequences like '4, 7, 10, 13, ?', where the number series itself specified the operation required to generate the next number.

7.4.1. Number comparison task

This task required patients to judge which of two numbers was numerically larger. There were twenty questions. Six involved comparisons between single-digit numbers (e.g., 9 and 2), two involved one versus two-digit numbers (e.g., 7 and 13) and eight involved comparisons between two-digit numbers (e.g., 10 and 16). In addition, two involved two-versus three-digit numbers (e.g., 105 and 89) and two involved comparisons between three-digit numbers (e.g., 948 and 199). All four patients were tested on this task. In every case, their performance was errorless suggesting that the patients' difficulties with the 'which number is nearest task' may have resulted from the fact that multiple comparisons and/or calculation were required. In addition, it is possible that the use of Arabic numbers rather than printed number words facilitated the patients' performance in this task, especially when the two numbers to be compared were different lengths (e.g., comparisons between two- and three-digit numbers).

7.4.2. Arithmetic questions

The four patients were asked to solve 108 calculations, written out on paper. They were allowed to write down their workings. There were 27 calculations for each of the mathematical operations (i.e., addition, subtraction, multiplication and division), presented in a mixed fashion. The particular operation required on each trial was indicated by the standard symbols, '+', '−', '×', and '÷'. The meaning of these symbols was explained to each patient prior to testing and it proved necessary to provide repeated reminders throughout the test. The sums involved one-, two- and three-digit operands (see Table 9). No control data is available for this task.

Incorrect responses were assigned to one of three categories: (1) 'no response' errors, (2) 'symbol comprehension' errors, which were responses that would have been correct if a different operation had been required (e.g., $21 - 9 = 30$; $14 \times 21 = 35$), and (3) 'other errors'. Some examples from this heterogeneous group are provided below. Table 9 shows the number of responses in each of these categories for each patient as a function of mathematical operation and operand size.

EK, GT and MK performed perfectly on the addition sums and PD only made a single error. The patients were also relatively good at subtraction. The less impaired patients made virtually no errors and the more severely impaired patients made a handful of errors on the multi-digit

Table 9
Calculation accuracy and error types

	No. of items[a]	EK				GT				PD				MK			
		Cor	NR	Symb	Other	Cor	NR	Symb	Other	Cor	NR	Symb	Other	Cor	NR	Symb	Other
Add																	
1 + 1	8 (7)	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
11 + 1	10 (9)	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
11 + 11	9	100	0	0	0	100	0	0	0	89	0	0	11	100	0	0	0
Overall	27 (25)	100	0	0	0	100	0	0	0	96	0	0	4	100	0	0	0
Subtract																	
1 − 1	4	100	0	0	0	100	0	0	0	100	0	0	0	100	0	0	0
11 − 1	10	90	0	0	10	100	0	0	0	90	10	0	0	90	0	10	0
11 − 11	10 (9)	90	0	0	10	100	0	0	0	78	0	0	22	60	10	0	30
111 − 111	3 (2)	100	0	0	0	100	0	0	0	50	0	0	50	100	0	0	0
Overall	27 (25)	93	0	0	7	100	0	0	0	84	4	0	12	81	4	4	11
Multiply																	
1 × 1	9 (8)	67	0	11	22	100	0	0	0	63	0	13	25	44	0	22	33
11 × 1	9	78	0	0	22	100	0	0	0	33	33	22	11	33	44	0	22
11 × 11	9 (8)	0	89	0	11	100	0	0	0	13	25	38	25	0	67	33	0
Overall	27 (25)	48	30	4	19	100	0	0	0	36	20	24	20	26	37	19	19
Divide																	
1/1	5	80	0	20	0	100	0	0	0	100	0	0	0	60	20	0	20
11/1	4	75	0	0	25	100	0	0	0	25	75	0	0	25	25	50	0
111/1	9 (8)	11	56	0	33	100	0	0	0	13	50	0	38	0	67	22	11
111/11	9 (8)	33	44	0	22	100	0	0	0	25	38	0	38	22	67	0	11
Overall	27 (25)	41	33	4	22	100	0	0	0	36	40	0	24	22	52	15	11
Grand mean	108 (100)	70	16	2	12	100	0	0	0	63	16	6	15	57	23	9	10

Figures refer to percentage of item presented. Cor: correct, NR: no response, Symb: symbol comprehension error.

[a] The numbers of items shown in parentheses refer PD, who was tested on fewer items due to time constraints.

problems. EK, MK and PD showed much poorer performance on multiplication and division questions, scoring an average of only 37 and 33%, respectively. In contrast, GT was able to carry out multiplication and division without difficulty. The patients' performance was strongly affected by the size of the operands.

The most frequent error was a failure to respond. The patients also repeatedly made errors that appeared to result from the selection of an inappropriate mathematical operation, suggesting that the patients may not have understood the meanings of the mathematic symbols. EK, MK and PD indicated that they understood the symbol '+', but did not comprehend '−', '×', or '÷'. GT comprehended all four symbols. The patients also made a considerable number of other calculation errors. Some of these errors were reminiscent of an impairment of arithmetic facts/rules; e.g., $2 \times 3 = 8$, $8 \times 1 = 16$ (McCloskey, 1992) and some may have resulted from failures to follow multi-digit procedures; e.g., $34 − 18 = 26$, rather than 16. The patients were able to read the Arabic numbers from 1 to 20 without error, suggesting that these errors did not result from an inability to recognise Arabic numbers.

7.4.3. 'Which number comes next' test

As the patients' failure to understand mathematical symbols appeared to contribute to their poor performance on the previous test, a calculation task was devised that avoided the use of such symbols. EK, GT, MK and twelve controls were tested on this 'which number comes next' task, which required the next number in a sequence to be calculated (for example, 4, 7, 10, 13, ?). The sequences themselves, presented in Arabic numerals, specified the operation that produced the answer. There were 40 sequences, divided equally between addition, subtraction, multiplication and division. The different types of sequences were blocked and the patients were told that the blocks required different operations. Four of the twenty addition and subtraction sequences were 'second order' as the amount that was added or subtracted was changed by a constant amount each time (e.g., 2, 3, 5, 8, ?).

The patients' performance on these sequences, shown in Table 10, was generally good. GT was again functioning at a particularly high level. EK showed some impairment on

Table 10
Performance of patients and controls on the number sequence test

	Maximum	EK	GT	MK	Control mean (range)
Addition	10	7	10	9	9.3 (7–10)
Subtraction	10	9	10	5[a]	9.2 (6–10)
Multiplication	10	9	10	3	7.4 (3–10)
Division	10	4[a]	10	4[a]	9.5 (7–10)
Higher order	4	1	4	4	3.0 (0–4)

[a] Denotes abnormal performance.

sequences involving division, and MK showed some impairment on subtraction, multiplication and division sequences. These deficits did not seem to be accountable by a failure to understand the task or the numbers involved—MK showed perfect performance on the higher order sequences, demonstrating that she was able to detect underlying patterns and make inferences about the next number. Instead, the impairments could be explained by a poor understanding of the multiplication and division procedures. The majority of EK's errors occurred because she divided the components of a multi-digit number correctly but failed to add the products of these calculations (for example, she recorded half of 16 as '53', apparently because she knew that half of 10 was 5, and that half of 6 was 3). MK's errors were omissions or resulted from the use of the wrong mathematical operation (addition for multiplication and division sequences).

7.4.4. Discussion

These results, taken together, provide evidence about which aspects of number knowledge remain intact in SD and which become degraded. The patients were able to make accurate judgements about numerical magnitude (e.g., in comparison tasks), could place number words in the correct order, were able to translate between Arabic numerals and number words and could perform naming and word-picture matching with dot pictures. In contrast, in line with previous reports (Cappelletti et al., 2001), their ability to perform calculations, particularly those requiring multiplication and division, was impaired. The patients showed a tendency to over-apply the addition procedure to other sums requiring different operations. Their knowledge of mathematical symbols was clearly degraded. This finding is of particular interest, as only one previous study has reported a specific difficulty with the comprehension of arithmetic signs (Ferro & Botelho, 1982). The patients also performed more poorly on arithmetic problems that involved multi-digit operands, and some of their errors on these questions implied a specific impairment of the procedures required for multi-digit numbers. In addition, they made some errors on very easy multiplication problems (e.g., $2 \times 3 = 8$, $8 \times 1 = 16$), which may have been indicative of an impairment of arithmetic facts (McCloskey, 1992).

This pattern of competencies and weaknesses is consistent with models of numerical cognition that postulate a distinction between semantic/parietal representations of numerical magnitude and verbal representations of number words and arithmetic facts (e.g., Dehaene, 1992; McCloskey, Caramazza, & Basili, 1985; Noel & Seron, 1993). In Dehaene 'triple code' model (1992), for example, there is an analogue representation of numerical magnitude which is separable from the verbal code for number words and the visual code for Arabic numerals. Certain types of numerical processing (e.g., transcoding, recall of arithmetic facts) are thought to draw heavily on the verbal and visual codes, whereas other tasks (e.g., magnitude comparison) are thought to be more independent of linguistic representations

(but see Noel & Seron, 1997). Brain regions within the inferior parietal lobe are thought to underpin knowledge of quantity (Dehaene & Cohen, 1995; Dehaene, Dehaene-Lambertz, & Cohen, 1998). Patients with SD should therefore have a good understanding of quantity, consistent with their intact performance on number comparison tasks. In contrast, the stable associations between verbal/Arabic numbers and representations of magnitude might become degraded in SD. Although the ability to comprehend and recall number words relative to other words could be partially protected because numbers have meaningful magnitude referents that are frequently encountered in the world, there is some evidence that transcoding tasks involving single-digit numbers can become impaired in cases with very severe SD (Knott, 1998).

It is also interesting to note that GT's number processing abilities were superior to those of EK and MK across a range of tasks. His particularly marked preservation of number knowledge may have been related to his extensive pre-morbid experience with number and calculation in his job as a college technician. Similarly, patient IH (Cappelletti et al., 2001) had exceptional number knowledge given his level of semantic impairment and had worked as a City banker. Therefore, pre-morbid experience of numbers may help to determine the extent to which they remain intact in this condition.

8. Reading aloud number and non-number words

If the patients' relatively good understanding of number helps to maintain the phonology of number words in ISR, a similar difference between number and non-number words might be expected to emerge in other apparently 'non-semantic' tasks requiring phonological production, for example, reading aloud. Some views about the translation from orthography to phonology suggest that semantic representations play an important role in reading aloud, especially for low frequency words with atypical spelling-to-sound correspondences (Plaut, McClelland, Seidenberg, & Patterson, 1996). SD patients make reading errors on such words, pronouncing them as if they had regular correspondences (PINT to rhyme with 'mint'): i.e., they demonstrate surface dyslexia (Graham, Hodges, & Patterson, 1994; Patterson & Hodges, 1992).

If this hypothesis is correct, SD patients might be expected to show preserved reading of irregular number words, as the correct pronunciations of these items should receive more support from semantics if knowledge of the number domain is relatively preserved. Cappelletti et al. reported that patient IH, who had good comprehension of number words, did make fewer errors on reading aloud number words, compared with other categories of words, even when the words were matched for frequency and regularity (Butterworth et al., 2001; Cappelletti et al., 2002). EK, GT and MK were tested using Cappelletti et al.'s materials,

in order to determine if they were also more accurate at reading number words compared with non-number words.

8.1. Method

The patients were asked to read 30 cardinal number words (the numbers from one to twenty, each tenth number—'thirty', 'forty', etc., and the words hundred, thousand and million). They were also asked to read 22 ordinal number words (first, second, etc., up to twentieth, and then the items hundredth and thousandth), and 18 'ambiguous' number related words that also had non-numerical meanings (add, minus, share, etc.). These 70 number words were categorised as having regular and irregular spelling patterns and were compared with 70 non-number words matched on frequency, spelling regularity and length.

8.2. Results

The percentage of number and non-number words read aloud correctly by the three patients is shown in Table 11. Butterworth et al. (2001) reported that control participants were errorless on this task. The patients were relatively good at reading both regular and irregular number words, and made a larger number of errors on the non-number words, although the differences were subtle compared with the dramatic difference shown by IH. The advantage for reading number words was statistically significant for both EK and GT (EK: $\chi^2(1) = 4.07$, $P < 0.05$; GT: $\chi^2(1) = 3.89$, $P < 0.05$). MK made a larger number of errors on both number and non-number words, and showed no significant differences between them ($\chi^2(1) < 1$). As anticipated, none of the patients showed an effect of regularity in their reading of number words ($\chi^2(1) < 1$). GT showed a marginally significant effect of regularity in his reading of non-number words (GT: $\chi^2(1) = 3.56$, $P = 0.059$), and the other two patients showed numerical advantages for regular over irregular non-number words that failed to reach significance (EK: $\chi^2(1) < 1$; MK: $\chi^2(1) = 1.96$, n.s.).

Table 11
Reading number and non-number words

	EK	GT	MK
Cardinal number words ($n = 30$)			
Regular ($n = 19$)	100	100	84
Irregular ($n = 11$)	100	100	91
Ordinal number words ($n = 22$)			
Regular ($n = 13$)	77	77	77
Irregular ($n = 9$)	78	100	67
Ambiguous number words ($n = 18$)			
Regular ($n = 14$)	93	100	57
Irregular ($n = 4$)	75	100	50
Non-number words ($n = 70$)			
Regular ($n = 46$)	78	91	74
Irregular ($n = 24$)	71	71	54

Note: Numbers are expressed as a percentage of items presented.

8.3. Discussion

EK and GT were better at reading numbers than non-number words matched for regularity, frequency and length, although MK did not show this effect, perhaps because her number comprehension was more impaired. The correct phonology was more likely to be produced for number than non-number words in both ISR and reading aloud; our hypothesis is that in both of these tasks, the phonology of the number words received stronger support from the semantic system.

9. Experiment 5: matching span for number and non-number words

The recall advantage for number words appears to be equivalent to the better recall of known than degraded words reported previously (e.g., Patterson et al., 1994). Therefore, these findings are consistent with the view that stable semantic representations play an important role in verbal STM. However, it is not clear *how* these representations act to support the phonology of items in STM, and there are several opposing viewpoints. The redintegration account (Gathercole et al., 2001; Hulme et al., 1991) suggests that long-term lexical representations are used to reconstruct the degraded phonological STM trace during recall. This process may be disrupted by the degradation of word knowledge in SD. In contrast, Patterson et al. (1994) accounted for the relationship between semantic knowledge and ISR within a parallel distributed processing (PDP) framework. According to this account, stable phonological and semantic patterns corresponding to familiar words and concepts conspire to maintain phonological coherence of items for ISR. In SD, this additional source of constraint may be lessened for semantically degraded words, making it more difficult for patients to retain these items as a coherent whole.

The redintegration account suggests that lexical representations exert their effects during overt recall, at a late stage in verbal STM. In contrast, the PDP account suggests that semantic representations are involved during all stages of ISR. Therefore, the two accounts predict rather different results in serial recognition tasks like matching span. In this task, two successive lists of items are read aloud to the participant, who is then required to make a same/different judgement. As this judgement does not require overt recall, the redintegration account predicts that lexical influences will be reduced or even abolished. In line with this suggestion, Knott et al. (1997) found no difference in matching span for known and degraded words in a patient with SD, and Gathercole et al. (2001) found a markedly reduced lexicality effect in matching span in normal children. This experiment examined whether the difference between numbers and words was limited to ISR, consistent with the redintegration account, or whether it extended to matching span, consistent with the PDP account.

The experiment compared two different types of changes in a matching span task—changes in item order and changes in phoneme order. Typical matching span tasks examine the ability to detect switches in the order of neighbouring items (e.g., 'her, sight, name, small' becomes 'her, name, sight, small'). We also examined the ability to detect switches in the onsets of neighbouring words (e.g., 'her, sight, name, small' versus 'her, night, same, small'). According to the PDP framework, semantic degradation leads to specific difficulty in maintaining the order of phonemes rather than whole items, explaining why SD patients make a large number of phoneme migration but not item order errors in recall (see Jefferies, Frankish et al., submitted-a). Therefore, SD patients might be more impaired at detecting changes in onsets than changes in item order, and the influence of semantic representations on matching span might prove more readily detectable using this method.

9.1. Method

Matching span performance was examined for the single-digit numbers and the frequency and imageability matched words used in Experiment 1 (see Appendix A). Two lists of digits or words were read aloud in succession, and the participants judged if the lists had been the same or different. Two types of changes could occur in the second list: (1) changes in item order, in which two neighbouring items were switched in order, and (2) changes in phoneme order, in which the onsets of two neighbouring words were exchanged. Examples of both types of changes were provided before the start of test.

The digit and word lists were yoked so that matched pairs of items occurred in the same serial positions. All lists were presented twice, allowing the two types of changes to be made on identical lists. Before the lists were constructed, the result of switching onsets between every possible combination of items was established. Some of these changes resulted in real words being produced ('nine, four' to 'fine, gnaw'), some resulted in non-words being produced ('seven, five' to 'feven, sive'), and some were impossible because the onsets for the two words were identical ('four, five'). Im-

possible changes were discarded, and changes that were impossible in one type of material were avoided for the yoked items in the other. Changes were selected from the remaining possibilities so that an equal number resulted in words and non-words for the two types of material. The same number of changes occurred at each serial position. The words that changed were placed in the lists first, and the remainder of the lists were constructed by selecting items at random without replacement and re-pooling the items as many times as required.

There were four conditions in this experiment, corresponding to the two types of changes for the two types of material, each with 24 trials. Changes occurred on half of these trials. The trials were presented in blocks of six trials, using a Latin square design. The patients and twelve matched controls were tested on five and seven item lists. The words were presented at a rate of one word per second, and the two lists were separated by a two second pause.

9.2. Results

The results for the patients and controls are shown in Table 12. The controls performed equally well on the matching span tasks for digits and words, whether changes were made to the order of the items (five item lists: $t(11) = 1.12$, n.s.; seven item lists: $t(11) < 1$), or to the order of the phonemes (five item lists: $t(11) < 1$; seven item lists: $t(11) = 1.12$, n.s). They were better able to detect changes in phoneme order than changes in item order, for both digits (five item lists: $t(11) = 3.35$, $P < 0.01$; seven item lists: $t(11) = 4.61$, $P < 0.001$), and words (five item lists: $t(11) = 4.17$, $P < 0.01$; seven item lists: $t(11) = 4.81$, $P < 0.001$).

The patients showed a rather different pattern of results. All three patients were unimpaired in detecting changes to the order of digits. They were also relatively good at detecting changes in the order of non-number items, although GT was impaired relative to controls at both list lengths. GT and MK were impaired at detecting changes to phoneme order in the digit items, at least in parts of the dataset. All three patients showed some impairment at detecting changes to the

Table 12
Matching span performance

List length	Matching condition	EK	GT	MK	Control mean (range)
Five items	Digits: item order	21	22	22	20.8 (18–23)
	Words: item order	18	17[a]	18	20.2 (18–23)
	Digits: phoneme order	24	19[a]	22	22.5 (21–24)
	Words: phoneme order	18[a]	16[a]	20	22.3 (20–24)
Seven items	Digits: item order	14	19	16	17.2 (12–22)
	Words: item order	14	10[a]	15	17.1 (11–20)
	Digits: phoneme order	21	21	16[a]	21.4 (18–24)
	Words: phoneme order	20	14[a]	15[a]	20.8 (17–24)

Maximum = 24, chance level = 12. Figures refer to number of correct trials.
   [a] Denotes abnormal performance.

order of phonemes in the non-number words. This was the only condition that was impaired in the mildest patient, EK.

EK showed no difference between digits and words when item order was changed (five and seven item lists: $\chi^2(1) < 1$), but she performed more poorly on words than digits when phoneme order was changed (five-item lists: $\chi^2(1) = 4.76$, $P < 0.05$; seven-item lists: $\chi^2(1) < 1$). GT performed better on the digits than the words when item order was changed (seven-item lists: $\chi^2(1) = 5.58$, $P < 0.05$; five-item lists: $\chi^2(1) = 2.19$, n.s.) and when phoneme order was changed (seven-item lists: $\chi^2(1) = 3.80$, $P = 0.051$; five-item lists: $\chi^2(1) < 1$). MK failed to show any significant differences between digits and words for either item order changes (five-item lists: $\chi^2(1) = 1.35$, n.s.; seven-item lists, $\chi^2(1) < 1$) or phoneme order changes (five- and seven-item lists: $\chi^2(1) < 1$).

9.3. Discussion

The controls' matching span performance did not differ between single-digit numbers and non-number words. In contrast, EK and GT showed a significant advantage for numbers over non-number words that mirrored the recall differences obtained in Experiment 1. MK was impaired on both the single-digit numbers and the non-number words and showed no significant differences between them. The broadly parallel findings obtained for EK and GT in recall and recognition tasks suggest that the role of semantics in verbal STM is not restricted to recall, in line with the predictions of the PDP account.

Although superior matching span performance for number words was not obtained in every patient on every list length, it appears that the redintegration account would always expect equivalent performance for these items; the fact that differences were obtained is therefore more consistent with the PDP view. It may be harder to observe the influence of semantic factors in matching span compared with ISR because the task is less demanding and less sensitive; it is possible to succeed at matching span, but not ISR, without a perfectly intact phonological representation.

10. General discussion

This series of experiments examined ISR of number and non-number words in patients with SD, in order to investigate the quantity and quality of span performance for these two types of materials. For every patient, the recall of single-digit numbers was normal whereas the recall of non-number words was impaired relative to controls, and this number advantage extended to lower frequency multi-digit numbers and words. In every experiment, the patients' recall revealed a relatively normal pattern of omission, order and intrusion errors on the number words but an abnormally large number of phonological errors on the non-number words. The difference between numbers

and non-number words remained substantial even when frequency, imageability, word length, set size and size of semantic category were matched across the two types of material.

The ISR differences between number and non-number words are reminiscent of those observed in SD patients previously for relatively well known versus semantically degraded words (Knott et al., 1997; Patterson et al., 1994), suggesting that a comprehension difference might underpin the better ISR for numbers. Naming and word-picture matching tests with dot and face part pictures supported this hypothesis, as the patients were able to perform these comprehension tasks for the numbers but not the matched non-number words. A similar association between comprehension and ISR was also observed within the domain of numbers, in a comparison of multi-digit and single-digit number words. Moreover, the patient with the largest comprehension impairment for the non-number words also exhibited the poorest ISR for these words.

There are at least two different accounts of the relationship between semantic (and/or other long-term representations) and ISR. The 'redintegration' account suggests first that it is stable phonological representations of familiar words, and not their meanings, that assist ISR, and secondly that the impact of these long-term representations on short-term memory occurs only at the time of recall when STM traces need reconstructing (Gathercole et al., 2001; Hulme et al., 1991). The alternative account suggests that, because of well-learned bi-directional connections between semantic and phonological representations, both types of stable knowledge constrain short-term phonological activation, and that this interactive support occurs during presentation and maintenance of information in STM as well as at the time of recall (Patterson et al., 1994). Verbal STM was better for number than non-number words in our SD patients not only in a recall paradigm but also in recognition, suggesting that the role of long-term representations in STM is not confined to the process of recall.

Why is number knowledge relatively preserved in SD? Cappelletti et al. (2001) suggested that straightforward explanations, like the high frequency and orderliness of numbers, could not account for IH's superior understanding of number because non-number words that shared these characteristics were not preserved. Similarly, the present article demonstrates that the selective preservation of ISR for numbers survives matching for frequency, imageability, word length, set size and open or closed semantic category. Another possibility is that numbers are preserved because they are important in everyday life; for example, they are essential for shopping and to use the telephone. Patients may retain an understanding of words and concepts that they encounter every day because of their preserved ability to form new episodic memories (Snowden, Griffiths, & Neary, 1994, 1996). In addition, number knowledge may be relatively preserved in SD because the cortical atrophy associated with this condition typically spares the inferior parietal lobe

thought to be crucial for the representation of numerical magnitude (see Dehaene et al., 1998, for a review).

Although SD patients' preserved representation of numerical magnitude may strengthen their comprehension and recall of number words relative to words in other categories, knowledge of number words ultimately degrades in SD. Our patients retained good understanding and recall of single-digit numbers but their knowledge and ISR of less frequent multi-digit numbers was somewhat impaired, consistent with the greater vulnerability of low frequency words and concepts in this condition (Funnell, 1995). Single-digit number words also refer to concepts that are easy to manipulate and acquired at a young age; factors which may have contributed to their better comprehension and recall. Despite their many advantages, even single-digit numbers can become degraded in the later stages of SD. Eventually, patient FM became impaired at naming and word-picture matching with the numerals 1–9 and at this same stage, she also began to make phonological errors during digit span (Knott, 1998). This association between knowledge and ISR within the domain of number words in the very late stages of SD lends further support to the view that semantics plays a crucial role in verbal STM.

Acknowledgements

Appendix A

The two word sets used in Experiment 1

	Digits	Frequency matched words	Frequency and imageability matched words
	One	All	Her
	Two	Well	Back
	Three	Though	Small
	Four	Lot	Name
	Five	Soon	Light
	Six	Road	Age
	Seven	Sorry	Council
	Eight	Worth	Health
	Nine	Bread	Sight
Mean syllable length	1.1	1.1	1.1
Mean frequency	748.9	786.8	799.7
Mean imageability	449.8	439.2[a]	458.0

[a] Score is unavailable for some items.

Appendix B

Numbers and matched words used in Experiment 2

	Numbers	Frequency matched	Frequency matched, high imageability
	Eleven	Article	Furniture
	Thirteen	Birthday	Tennis
	Seventeen	Definite	Envelope
	Nineteen	Trading	Dusty
	Seventy	Notably	Pollution
	Eighty	Shiny	Cigar
	Ninety	Applause	Cement
	Billion	Gesture	Novel
	Trillion	Blandly	Madman
Mean syllable length	2.3	2.3	2.3
Mean frequency	23.4	26.5	24.3
Mean imageability	Not available	Not available	564.4

Appendix C

Number and face-part words used in Experiment 3

	Numbers	Frequency matched words
	Eight	Mouth
	Nine	Neck
	Ten	Hair
	Eleven	Tongue
	Twelve	Nose
	Thirteen	Brow
	Fourteen	Beard
	Sixteen	Chin
	Eighteen	Forehead
	Nineteen	Fringe
	Sixty	Cheek
	Seventy	Parting
Mean syllable length	1.8	1.25
Mean frequency	61.1	62.5
Mean imageability	Not available	599.7[a]

[a] Score is unavailable for some items.

References

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX Lexical database [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Bourassa, D. C., & Besner, D. (1994). Beyond the articulatory loop—A semantic contribution to serial order recall of subspan lists. *Psychonomic Bulletin and Review*, *1*, 122–125.

Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., & Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia*, *38*, 1207–1215.

Bozeat, S., Lambon Ralph, M. A., Patterson, K., & Hodges, J. R. (2002). The influence of personal familiarity and context on object use in semantic dementia. *Neurocase*, *8*, 127–134.

Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology*, *26*, 467–482.

Butterworth, B., Cappelletti, M., & Kopelman, M. (2001). Category-specificity in reading and writing: the case of number words. *Nature Neuroscience*, *4*, 784–786.

Cappelletti, M., Butterworth, B., & Kopelman, M. (2001). Spared numerical abilities in a case of semantic dementia. *Neuropsychologia*, *39*, 1224–1239.

Cappelletti, M., Kopelman, M., & Butterworth, B. (2002). Why semantic dementia drives you to the dogs (but not to the horses): A theoretical account. *Cognitive Neuropsychology*, *19*, 483–503.

Caza, N., Belleville, S., & Gilbert, B. (2002). How loss of meaning with preservation of phonological word form affects immediate serial recall performance: A linguistic account. *Neurocase*, *8*, 255–273.

Cipolotti, L., & Butterworth, B. (1995). Towards a multiroute model of number processing: impaired transcoding with impaired calculation skills. *Journal of Experimental Psychology: General*, *124*, 375–390.

Cipolotti, L., Butterworth, B., & Denes, G. (1991). A specific deficit for numbers in a case of dense acalculia. *Brain*, *114*, 2619–2637.

Cohen, L., Verstichel, P., & Dehaene, S. (1997). Neologistic jargon sparing numbers: a category-specific phonological impairment. *Cognitive Neuropsychology*, *14*, 1029–1061.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33A*, 497–505.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, *55*, 432–439.

Crutch, S. J., & Warrington, E. K. (2002). Preserved calculation skills in a case of semantic dementia. *Cortex*, *38*, 389–399.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*, 1–42.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of numerical processing. *Mathematical Cognition*, *1*, 83–120.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, *33*, 219–250.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neuroscience*, *21*, 355–361.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43*, 1–29.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284*, 970–974.

Dehaene, S., Tzourio, N., Frak, V., Raynaud, L., Cohen, L., & Mehler, J. et al., (1996). Cerebral activations during number multiplication and comparison: A PET study. *Neuropsychologia*, *34*, 1097–1106.

Delazer, M., & Benke, T. (1997). Arithmetic facts without meaning. *Cortex*, *33*, 697–710.

Delazer, M., & Butterworth, B. (1997). A dissociation of number meanings. *Cognitive Neuropsychology*, *14*, 613–636.

Deloche, G., & Seron, X. (1982). From one to 1: An analysis of a transcoding process by means of neuropsychological data. *Cognition*, *12*, 119–149.

Diesfeldt, H. F. A. (1993). Progressive decline of semantic memory with preservation of number processing and calculation. *Behavioural Neurology*, *6*, 239–242.

Evans, J. J., Heggs, A. J., Antoun, N., & Hodges, J. R. (1995). Progressive prosopagnosia associated with selective right temporal-lobe atrophy: A new syndrome. *Brain*, *118*, 1–13.

Ferro, J. M., & Botelho, S. (1982). Alexia for arithmetical signs: A cause of disturbed calculation. *Cortex*, *16*, 175–180.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Forde, E. M. E., & Humphreys, G. W. (2002). The role of semantic knowledge in short-term memory. *Neurocase*, *8*, 13–27.

Funnell, E. (1995). Objects and properties: A study of the breakdown of semantic memory. *Memory*, *3*, 497–518.

Funnell, E. (1996). Response biases in oral reading: An account of the co-occurrence of surface dyslexia and semantic dementia. *Quarterly Journal of Experimental Psychology*, *49A*, 417–446.

Galton, C. J., Patterson, K., Graham, K., Lambon Ralph, M. A., Williams, G., & Antoun, N. et al., (2001). Differing patterns of temporal atrophy in Alzheimer's disease and semantic dementia. *Neurology*, *57*, 216–225.

Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology*, *54A*, 1–30.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition, 87*, B87–B95.

Graham, K. S., Hodges, J. R., & Patterson, K. (1994). The relationship between comprehension and oral reading in progressive fluent aphasia. *Neuropsychologia*, *32*, 299–316.

Graham, N. L., Patterson, K., & Hodges, J. R. (2001). The emergence of jargon in progressive fluent dysgraphia: The widening gap between target and response. *Cognitive Neuropsychology*, *18*, 343–361.

Hodges, J. R., Patterson, K., Oxbury, S., & Funnell, E. (1992). Semantic dementia: Progressive fluent aphasia with temporal-lobe atrophy. *Brain*, *115*, 1783–1806.

Howard, D., Patterson, K., 1992. Pyramids and Palm Trees: A test of semantic access from pictures and words. Bury St. Edmunds, Suffolk: Thames Valley Test Company.

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, *30*, 685–701.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *23*, 1217–1232.

Jefferies, E., Frankish, C. R., & Lambon Ralph, M. A. (submitted-a). Lexical and semantic binding in short-term memory: Evidence from normal recall and semantic dementia.

Jefferies, E., Jones, R., Bateman, D., & Lambon Ralph, M. A. (submitted-b). A semantic contribution to nonword recall? Evidence for intact phonological processes in semantic dementia.

Jefferies, E., Jones, R., Bateman, D., & Lambon Ralph, M. A. (submitted-c). When does word meaning affect immediate serial recall in semantic dementia?

Knott, R. (1998). The impact of lexical and semantic impairment on short-term verbal recall. Unpublished thesis, University of Cambridge, Cambridge.

Knott, R., Patterson, K., & Hodges, J. R. (1997). Lexical and semantic binding effects in short-term memory: Evidence from semantic dementia. *Cognitive Neuropsychology*, *14*, 1165–1216.

Knott, R., Patterson, K., & Hodges, J. R. (2000). The role of speech production in auditory-verbal short-term memory: Evidence from progressive fluent aphasia. *Neuropsychologia*, *38*, 125–142.

Lambon Ralph, M. A., & Howard, D. (2000). Gogi aphasia or semantic dementia? Simulating and assessing poor verbal comprehension in a case of progressive fluent aphasia. *Cognitive Neuropsychology*, *17*, 437–465.

Lezak, M. (1976). Neuropsychological Assessment. New York: Oxford University Press.

Martin, R. C., & Lesch, M. (1996). Associations and dissociations between language impairment and list recall: Implications for models of STM. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 149–178). East Sussex, UK: Psychology Press.

Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, *14*, 641–682.

Martin, R. C., Shelton, J., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language*, *33*, 83–111.

McCarthy, R. A., & Warrington, E. K. (1987). The double dissociation of short-term memory for lists and sentences: Evidence from aphasia. *Brain*, *110*, 1545–1563.

McCarthy, R. A., & Warrington, E. K. (2001). Repeating without semantics: Surface dysphasia? *Neurocase*, *7*, 77–87.

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, *44*, 107–157.

McCloskey, M., Caramazza, A., & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition*, *4*, 171–196.

Mummery, C. J., Patterson, K., Price, C. J., Ashburner, J., Frackowiak, R. S. J., & Hodges, J. R. (2000). A voxel-based morphometry study of semantic dementia: Relationship between temporal-lobe atrophy and semantic memory. *Annals of Neurology*, *47*, 36–45.

Noel, M. P., & Seron, X. (1993). Arabic number reading deficit: A single case study or when 236 is read (2306) and judged superior to 1258. *Cognitive Neuropsychology*, *10*, 317–339.

Noel, M. P., & Seron, X. (1997). On the existence of intermediate representations. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *23*, 697–720.

Patterson, K., Graham, N., & Hodges, J. R. (1994). The impact of semantic memory loss on phonological representations. *Journal of Cognitive Neuroscience*, *6*, 57–69.

Patterson, K., & Hodges, J. R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia*, *30*, 1025–1040.

Plaut, D., McClelland, J. L., Seidenberg, M., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56–115.

Poirier, M., & Saint Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *Quarterly Journal of Experimental Psychology*, *48A*, 384–404.

Raven, J. C. (1962). *Coloured progressive matrices sets A, AB, B*. London: Lewis.

Snowden, J. S., Griffiths, H., & Neary, D. (1994). Semantic dementia: Autobiographical contribution to preservation of meaning. *Cognitive Neuropsychology*, *11*, 265–288.

Snowden, J. S., Griffiths, H. L., & Neary, D. (1996). Semantic-episodic memory interactions in semantic dementia: Implications for retrograde memory function. *Cognitive Neuropsychology*, *13*, 1101–1137.

Snowden, J. S., Neary, D., & Mann, D. M. A. (Eds.) (1996). *Frontotemporal lobar degeneration: Frontotemporal dementia, progressive aphasia, semantic dementia*. London: Churchill Livingstone.

Stanescu-Cosson, R., Pinel, P., van de Moortele, P., Le Bihan, D., Cohen, L., & Dehaene, S. (2000). Understanding dissociations in dyscalculia: A brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. *Brain*, *123*, 2240–2255.

Thioux, M., Pillon, A., Samson, D., de Partz, M. P., Noel, M. P., & Seron, X. (1998). The isolation of numerals at the semantic level. *Neurocase*, *4*, 371–389.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 145–152.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*, 1256–1271.

Warrington, E. K. (1975). The selective impairment of semantic memory. *The Quarterly Journal of Experimental Psychology*, *27*, 635–657.

Warrington, E. K. (1982). The fractionation of arithmetical skills: A single case study. *Quarterly Journal of Experimental Psychology*, *34A*, 31–51.

Warrington, E. K., & James, M. (1991). *The Visual Object and Space Perception battery*. Bury St. Edmunds, Suffolk: Thames Valley Test Company.

Wechsler, D. (1987). *Wechsler memory scale—revised* (*WMS-R*). New York: Psychological Corporation.

Wechsler, D. (1997). *Wechsler memory scale—third edition* (*WMS-III*). London: The Psychological Corporation.